The Life Cycle Process of Science Data Generation and Archive

The Role of the Active Archive

Second Seeds Workshop
San Diego

June 18, 2002

Steve Kempler
GES DAAC Manager
steve.kempler@gsfc.nasa.gov
301-614-5765

Scope of Active Archives

- *Active Archives are responsible for the execution, engineering support, data support, and science support of mission oriented data management functions (data ingest, archive, production, distribution, management, and user services).*
- *Active Archives are the keepers of data and information generated by data sources, derived by data processing, and created by data providers.*
- *Active Archives are staffed by experts in discipline specific science fields, computer engineering and technology, production operations, and user services.*
- *Active Archives facilitate science by developing value added products and services in response to the evolving needs of the science community and advances in data and information system technologies*
- *Active Archives remain active from mission preparedness and system development to mission completion and data access inactivity.*

Definition: An Active Archive

- **Active: Constantly engaged in action**
- **Archive: A place where records or documents are kept**

Thus, an Active Archive is the place where data and supporting information holdings are regularly ingested, stored, accessed and distributed.

V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V^V

In terms of facilitating the long term safeguard of NASA Earth science data, this implies that Active Archives must have:

- All available <u>Supporting Information ingested</u> and stored for all data, to maintain a thorough history of the data (**Information collection**)
- All data and supporting information <u>stored</u>, periodically checked for integrity, and backed up (**Information Stewardship**)
- All data and supporting information in a form that can be <u>accessible and usable</u> using current technologies, in preparation for LTA. (**Information migration**)

Active Archives must have the ability to prepare for stewarding data, steward data, and provide information for future stewardship of data, through the active period of the data lifecycle.

Active Archive Data Life Cycle Process Preparations

- Develop early relationship with data providers to understand mission goals and acquire data product documentation with generated data
- Ensure resident expertise in science disciplines, data management, and information systems
- Completely understand user community ‘model(s)’
- Develop tools that enhance the accessibility, usabilty, and value of the data
- Develop a philosophical, technical and operational relationship with the LTA sponsor

Level of User Services Needed for Life Cycle Process

Marked in Red

SERVICES (Levels)	LEVELS OF SERVICE (higher levels include lower levels, but not lowest)	INFORMATION REQUIRED OF THE DATA PRODUCER	SKILLS/KNOWLEDGE/EXPERIENCE NEEDED
USER SERVICES			
Lowest	No Service	None	
Low	Telephone, E-mail	Science data point of contact; Data formats and organization	Keep records/metrics; User services personnel
Medium	On-line help; Perhaps FAQ; Data Support: Knowledge of data, able to solve technical, data, data structure questions	Algorithm Information (ATBDs); Data formats and organization; Data system specification/performance	Computer engineers; Scientists; Software developers; Expertise in data management systems, and data structures
High	Outreach; Conferences; Papers and Presentations	Science peer review of material when appropriate	Outreach personnel; Poster/ presentation developers
Highest	Educational Material; Lectures and Seminars	Science peer review of material	Educators

Levels of Service for Products Support Needed for Life Cycle Process

Marked in Red

SERVICES (Levels)	LEVELS OF SERVICE (higher levels include lower levels, but not lowest)	INFORMATION REQUIRED OF THE DATA PRODUCER	KNOWLEDGE / EXPERIENCE/ SKILLS NEE DED
PRODUCT SUPPORT			
Lowest	No Support; Unmaintained basic documentation	None	
Low	Documentation; Webpages with hyperlinks to metadata; Readme's; DIFs	Description of data files, collection, granule, and product level metadata, etc.; Product information; Product descriptions; Guide documents; Browse products; Product volume, restrictions, and quality; Science point of contact; Data format and access definition	Scientists; Software developers; Webmaster; Technical writers; Data specilaists
Medium	Science Support: Interpret data products; Develop or make available, tools to facilitate product usability and accessibility: e.g., format conversions, subsetting, data mining, additional distribution mechanisms	Science team consultation	Computer engineers
High	Development of value added products to meet the needs of users; Field campaign support to supplement current products; Product regeneration	Product generation software (i.e., PGEs) and complete documentation; PGE System requirements; PGE execution procedures, conditions, error handling, and initiation triggers	Science integrators
Highest	Center of Excellence concepts	Science peer review	Computer architects

Levels of Service of Long Term Data Stewardship Needed for Life Cycle Process

Marked in Red

SERVICES (Levels)	LEVELS OFSERVICE (higher levels include lower levels, but not lowest)	INFORMATION REQUIRED OF THE DATA PRODUCER	KNOWLEDGE/ EXPERIENCE/ SKILLS NEEDED
Long Term Data Stewardship			
Lowest	Keep data safe offline; Improve data access only when necessary; Media migration when needed; All required security	Data Security documentation	
Low	Keep data safe on or nearline; Evolve data access with affordable changes; Media migration when needed; Technology (e.g., data formats, media) and infrastructure (e.g., system) analysis and migration when needed; All required security	Algorithm Information (ATBDs); Data formats and organization; Data system specification/performance; Data Security documentation	Computer engineers; Data specialists
Medium	Evolve data access with user needs and/or to meet standards; Technology analysis and migration when appropriate and/or to meet standards; Infrastructure analysis and migration to meet industry improvements		
High	Ensure data safety further with offsite backup; Engage an advisory group to guide the data center in long term data stewardship planning		Scientists
Highest	Further ensure data stewardship through membership/participation in many standards committees, technology groups, etc.		